

**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE**

In re U.S. Patent Application of )  
)  
NISHIKAWA et al. )  
)  
Application Number: To be assigned )  
)  
Filed: Concurrently herewith )  
)  
For: AMINO ACID FRAME INDICATION SYSTEM, )  
METHOD FOR AMINO ACID FRAME INDICATION, )  
AND RECORDING MEDIUM )

Honorable Assistant Commissioner  
for Patents  
Washington, D.C. 20231

#3  
9-16-02

10857 U.S. PTO  
09/940664  
08/29/01

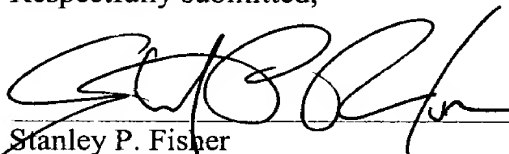
**REQUEST FOR PRIORITY  
UNDER 35 U.S.C. § 119  
AND THE INTERNATIONAL CONVENTION**

Sir:

In the matter of the above-captioned application for a United States patent, notice is hereby given that the Applicant claims the priority date of October 25, 2000, the filing date of the corresponding Japanese patent application 2000-325403.

The certified copy of corresponding Japanese patent application 2000-325403 is submitted herewith. Acknowledgment of receipt of the certified copy is respectfully requested in due course.

Respectfully submitted,

  
Stanley P. Fisher  
Registration Number 24,344

**REED SMITH HAZEL & THOMAS LLP**  
3110 Fairview Park Drive  
Suite 1400  
Falls Church, Virginia 22042  
(703) 641-4200  
August 29, 2001

**JUAN CARLOS A. MARQUEZ**  
Registration No. 34,072

( Translation )

PATENT OFFICE  
JAPANESE GOVERNMENT

This is to certify that the annexed is a true copy of  
the following application as filed with this Office.

Date of Application: October 25, 2000

Application Number: Japanese Patent Application  
No. 2000-325403

Applicant(s): HITACHI, LTD.  
HELIX RESEARCH INSTITUTE



June 1, 2001

Commissioner,  
Patent Office

Kozo OIKAWA (seal)

Certificate No. 2001-3051657

日 本 国 特 許 庁  
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office

出 願 年 月 日

Date of Application:

2000年10月25日

出 願 番 号

Application Number:

特願2000-325403

出 願 人

Applicant(s):

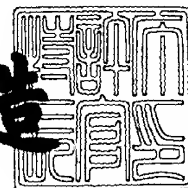
株式会社日立製作所  
株式会社ヘリックス研究所



2001年 6月 1日

特 許 庁 長 官  
Commissioner,  
Japan Patent Office

及 川 耕 造



出証番号 出証特2001-3051657

【書類名】 特許願

【整理番号】 H001696

【提出日】 平成12年10月25日

【あて先】 特許庁長官 殿

【国際特許分類】 C12Q 1/68

【発明者】

【住所又は居所】 東京都国分寺市東恋ヶ窪一丁目280番地 株式会社  
日立製作所 中央研究所内

【氏名】 西川 哲夫

【発明者】

【住所又は居所】 東京都千代田区神田駿河台四丁目6番地 株式会社 日  
立製作所 ライフサイエンス推進事業部内

【氏名】 村上 勝彦

【発明者】

【住所又は居所】 茨城県稲敷郡阿見町大室511-12

【氏名】 磯貝 隆夫

【発明者】

【住所又は居所】 東京都東大和市桜が丘3-44-14-9-204

【氏名】 永井 啓一

【発明者】

【住所又は居所】 大阪府豊中市曾根東町2-10-2-2-236

【氏名】 林 浩司

【発明者】

【住所又は居所】 千葉県木更津市請西1-25-13-103

【氏名】 入江 亮太郎

【発明者】

【住所又は居所】 千葉県木更津市朝日3-1-10-102

【氏名】 大槻 哲嗣

【特許出願人】

【識別番号】 000005108

【氏名又は名称】 株式会社 日立製作所

【特許出願人】

【識別番号】 597059742

【氏名又は名称】 株式会社 ヘリックス研究所

【代理人】

【識別番号】 100091096

【弁理士】

【氏名又は名称】 平木 祐輔

【手数料の表示】

【予納台帳番号】 015244

【納付金額】 21,000円

【その他】 国等の委託研究の成果に係る特許出願（平成 1 1 年度新エネルギー・産業技術総合開発機構（再）委託研究、産業活力再生特別措置法第 3 0 条の適用を受けるもの）

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9003115

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 アミノ酸フレーム表示システム、アミノ酸フレーム表示方法及び記録媒体

【特許請求の範囲】

【請求項 1】 cDNA 配列を入力する入力手段と、該入力された cDNA 配列を 1 文字ずつずらして翻訳して 3 つのアミノ酸フレームを得る翻訳手段と、前記入力された cDNA 配列とデータベース中の DNA 配列又はアミノ酸配列とのアラインメントを生成し、そのアラインメントから類似性情報に基づいて前記入力された cDNA 配列から翻訳されるアミノ酸配列を決定するアラインメント手段と、前記 3 つのアミノ酸フレーム上にまたがって前記アラインメント手段によって決定されたアミノ酸配列の領域を線分で表示する表示手段とを備えることを特徴とするアミノ酸フレーム表示システム。

【請求項 2】 前記アラインメント手段は、前記翻訳手段によって前記入力された cDNA 配列をあらかじめ 3 種類又は 6 種類の読み枠において翻訳したアミノ酸配列とデータベース中のアミノ酸配列との間のアラインメントから決定することを特徴とする請求項 1 記載のアミノ酸フレーム表示システム。

【請求項 3】 前記アラインメント手段は、前記入力された cDNA 配列中のコドン内ギャップを考慮にいて決定することを特徴とする請求項 1 記載のアミノ酸フレーム表示システム。

【請求項 4】 前記アラインメント手段は、前記入力された cDNA 配列とデータベース中の DNA 配列との間でそれぞれの DNA 配列中のコドン内ギャップを考慮にいて、それぞれの配列の翻訳されたアミノ酸配列間アラインメントから決定することを特徴とする請求項 1 記載のアミノ酸フレーム表示システム。

【請求項 5】 前記表示手段は、前記 3 つのアミノ酸フレームと共に、前記生成されたアラインメントを線状に表示することを特徴とする請求項 1 記載のアミノ酸フレーム表示システム。

【請求項 6】 前記表示手段は、前記線状に表示されたアラインメント中に、DNA 配列中の挿入又は欠失位置を表示することを特徴とする請求項 5 記載のアミノ酸フレーム表示システム。

【請求項 7】 前記表示手段は、前記線状に表示されたアラインメント中に、アラインメントの局所的な一致度を色によって表示することを特徴とする請求項 5 記載のアミノ酸フレーム表示システム。

【請求項 8】 前記表示手段は、前記 3 つのアミノ酸フレームと共に、前記生成された cDNA 配列とデータベース中の DNA 配列又はアミノ酸配列とのアラインメントをテキストで表示することを特徴とする請求項 1 記載のアミノ酸フレーム表示システム。

【請求項 9】 cDNA 配列を入力する入力手段と、該入力された cDNA 配列を 1 文字ずつずらして翻訳して 3 つのアミノ酸フレームを得る翻訳手段と、該 3 つのアミノ酸フレームにおける各開始コドン及び終始コドンを予測するコドン予測手段と、前記 3 つのアミノ酸フレーム上に前記開始コドン及び終始コドンの位置を表示すると共に、開始コドン位置に開始コドンらしさを表現する量又は記号を表示する表示手段とを備えることを特徴とするアミノ酸フレーム表示システム。

【請求項 10】 cDNA 配列を入力する入力手段と、該入力された cDNA 配列を 1 文字ずつずらして翻訳して 3 つのアミノ酸フレームを得る翻訳手段と、該 3 つのアミノ酸フレームにおける各開始コドン及び終始コドンを予測するコドン予測手段と、前記 3 つのアミノ酸フレームそれぞれにおけるコーディング領域らしさを表すコーディングポテンシャルを計算するコーディングポテンシャル計算手段と、前記 3 つのアミノ酸フレーム上に前記開始コドン及び終始コドンの位置を表示すると共に、各フレーム上又は別ウィンドーに、前記 3 つのアミノ酸フレームのコーディングポテンシャルを表示する表示手段とを備えることを特徴とするアミノ酸フレーム表示システム。

【請求項 11】 前記入力された cDNA 配列を編集して、編集後の cDNA 配列を改めて前記入力された cDNA 配列とする編集手段を備えることを特徴とする請求項 1、9 又は 10 記載のアミノ酸フレーム表示システム。

【請求項 12】 前記編集手段は、アラインメントのテキスト表示中において編集が実施可能であることを特徴とする請求項 11 記載のアミノ酸フレーム表示システム。

【請求項 1 3】 cDNA 配列を入力する入力ステップと、該入力された cDNA 配列を 1 文字ずつずらして翻訳して 3 つのアミノ酸フレームを得る翻訳ステップと、前記入力された cDNA 配列とデータベース中の DNA 配列又はアミノ酸配列とのアラインメントを生成し、そのアラインメントから類似性情報に基づいて前記入力された cDNA 配列から翻訳されるアミノ酸配列を決定するアラインメントステップと、前記 3 つのアミノ酸フレーム上にまたがって前記アラインメントステップによって決定されたアミノ酸配列の領域を線分で表示する表示ステップとを備えることを特徴とするアミノ酸フレーム表示方法。

【請求項 1 4】 cDNA 配列を入力する入力ステップと、該入力された cDNA 配列を 1 文字ずつずらして翻訳して 3 つのアミノ酸フレームを得る翻訳ステップと、該 3 つのアミノ酸フレームにおける各開始コドン及び終始コドンを予測するコドン予測ステップと、前記 3 つのアミノ酸フレーム上に前記開始コドン及び終始コドンの位置を表示すると共に、開始コドン位置に開始コドンらしさを表現する量又は記号を表示する表示ステップとを備えることを特徴とするアミノ酸フレーム表示方法。

【請求項 1 5】 cDNA 配列を入力する入力ステップと、該入力された cDNA 配列を 1 文字ずつずらして翻訳して 3 つのアミノ酸フレームを得る翻訳ステップと、該 3 つのアミノ酸フレームにおける各開始コドン及び終始コドンを予測するコドン予測ステップと、前記 3 つのアミノ酸フレームそれぞれにおけるコーディング領域らしさを表すコーディングポテンシャルを計算するコーディングポテンシャル計算ステップと、前記 3 つのアミノ酸フレーム上に前記開始コドン及び終始コドンの位置を表示すると共に、各フレーム上又は別ウィンドーに、前記 3 つのアミノ酸フレームのコーディングポテンシャルを表示する表示ステップとを備えることを特徴とするアミノ酸フレーム表示方法。

【請求項 1 6】 コンピュータを、cDNA 配列を入力する入力手段と、該入力された cDNA 配列を 1 文字ずつずらして翻訳して 3 つのアミノ酸フレームを得る翻訳手段と、前記入力された cDNA 配列とデータベース中の DNA 配列又はアミノ酸配列とのアラインメントを生成し、そのアラインメントから類似性情報に基づいて前記入力された cDNA 配列から翻訳されるアミノ酸配列を決定



するアラインメント手段と、前記 3 つのアミノ酸フレーム上にまたがって前記アラインメント手段によって決定されたアミノ酸配列の領域を線分で表示する表示手段とを備えるアミノ酸フレーム表示システムとして機能させるためのプログラムを記録したことを特徴とするコンピュータ読み取り可能な記録媒体。

【請求項 1 7】 コンピュータを、cDNA 配列を入力する入力手段と、該入力された cDNA 配列を 1 文字ずつずらして翻訳して 3 つのアミノ酸フレームを得る翻訳手段と、該 3 つのアミノ酸フレームにおける各開始コドン及び終始コドンを予測するコドン予測手段と、前記 3 つのアミノ酸フレーム上に前記開始コドン及び終始コドンの位置を表示すると共に、開始コドン位置に開始コドンらしさを表現する量又は記号を表示する表示手段とを備えるアミノ酸フレーム表示システムとして機能させるためのプログラムを記録したことを特徴とするコンピュータ読み取り可能な記録媒体。

【請求項 1 8】 コンピュータを、cDNA 配列を入力する入力手段と、該入力された cDNA 配列を 1 文字ずつずらして翻訳して 3 つのアミノ酸フレームを得る翻訳手段と、該 3 つのアミノ酸フレームにおける各開始コドン及び終始コドンを予測するコドン予測手段と、前記 3 つのアミノ酸フレームそれぞれにおけるコーディング領域らしさを表すコーディングポテンシャルを計算するコーディングポテンシャル計算手段と、前記 3 つのアミノ酸フレーム上に前記開始コドン及び終始コドンの位置を表示すると共に、各フレーム上又は別ウィンドーに、前記 3 つのアミノ酸フレームのコーディングポテンシャルを表示する表示手段とを備えるアミノ酸フレーム表示システムとして機能させるためのプログラムを記録したことを特徴とするコンピュータ読み取り可能な記録媒体。

# 【発明の詳細な説明】

## 【0 0 0 1】

### 【発明の属する技術分野】

本発明は遺伝子配列の情報解析に係わり、遺伝子配列がコードするアミノ酸配列を同定するためのアミノ酸フレーム表示システム、アミノ酸フレーム表示方法及び記録媒体に関する。

## 【0 0 0 2】

## 【従来の技術】

ヒトゲノムプロジェクトの進展（2000年6月にドラフトシーケンス完了）に伴い、配列決定のスループットが増大すると共に、遺伝子配列データベースの規模が急速に拡大している。大量に登録されているEST配列（遺伝子配列の部分配列）やドラフト配列（ゲノム配列の完了前の配列精度の低い配列）は、スループットを重視して集められた配列であり、配列精度はあまり高くない（EST配列は約3%のエラーがあるといわれている）。これらの配列から、できるだけ高精度のアミノ酸配列情報を抽出することが求められている。従来、cDNA配列からのアミノ酸配列情報の抽出には、アミノ酸フレーム表示を用いることが一般に行われていた（ORF Finder、<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>）。

## 【0003】

アミノ酸フレーム表示は、cDNA配列の5'端から1文字ずつずらして翻訳して得られる3つのアミノ酸配列を3つの線分として表示する。逆相補鎖も考慮に入れる場合は、全体で6つのアミノ酸配列を6つの線分として表示する。この線分上に開始コドンと終始コドンの位置を区別して表示し、開始コドンから始まり終始コドンで終了する線分を同定する。

## 【0004】

この線分を可能なオープンリーディングフレーム（ORF）として同定し、その中でもっとも長いORFを、このcDNAから抽出されたアミノ酸配列として同定していた。cDNA配列にフレームシフトエラーが存在する場合、アミノ酸フレーム表示上ではORFが分断され二つのフレームにまたがって表示される。しかも、分断されたORFの境界は明確ではなく、一般的には数十塩基の誤差で同定できるだけである。従って、cDNA配列にフレームシフトエラーが存在する場合には、従来、既知のアミノ酸配列との類似性情報を用いてフレームシフトエラーを同定することが行われてきた。cDNA配列をアミノ酸配列と比較する最も一般的なプログラムは、米国The National Center for Biotechnology Information（NCBI）で開発されたBLASTX（Altschul, S.F., et al., Basic local alignment search tool, J. Mol. Biol., 215(3), 403, 1990）である。BLASTXは、与えられたcDNA配列を6通りのアミノ酸配列に翻訳して（6フレーム）

、データベース中のアミノ酸配列との間で類似性比較を行い、アミノ酸配列間アラインメントを結果として出力する。cDNA配列にフレームシフトエラーが一つ存在する場合には、本来得られるべき一つのアラインメントが二つのアラインメントに分断されることになる。全体の類似度が高い場合には、分断されたアラインメントから元のアラインメントを再構成し、フレームシフト部位を同定することも、手間はかかるが可能である。しかし、全体の類似度が低い場合には、分断されたアラインメントから元のアラインメントを再構成しフレームシフト部位を同定することは困難である。cDNA配列をアミノ酸配列と比較する方法として、フレームシフトエラーを考慮した上で、アラインメントを得る方法が公開されている（特開平10-5000号公報）。この方法を用いれば、フレームシフトエラーが存在する場合にも一つのアラインメントが得られ、フレームシフト部位を同定することが可能である。しかし、この方法を用いる場合にも、全体の類似度が低い場合には、得られるアミノ酸配列の信頼性を評価することは困難である。このように、cDNA配列からアミノ酸配列を抽出するためには、アミノ酸フレームを用いる方法と、既知のアミノ酸配列との類似性情報を用いる方法があるが、cDNA配列にフレームシフトエラーが存在する場合にも信頼度の高いアミノ酸配列を抽出するためには、それぞれの方法のみの適用では不十分である。

## 【0005】

## 【発明が解決しようとする課題】

本発明が解決しようとする課題は、cDNA配列にフレームシフトエラーが存在する場合にも、信頼度の高いアミノ酸配列をcDNA配列から効率良く抽出することができるアミノ酸フレーム表示システム、アミノ酸フレーム表示方法及び記録媒体を提供することである。

## 【0006】

## 【課題を解決するための手段】

本発明では、着目する遺伝子配列に対して、配列の統計的解析、及び既知アミノ酸配列との類似性解析を行い、その結果をアミノ酸フレーム上に統合的に表示することによって、配列中のフレームシフトエラーの同定と編集処理を精度良くかつ効率良く実施可能にしている。

【 0 0 0 7 】

このために、cDNA配列に対して、以下の処理ステップ

- (1) 開始コドン予測プログラムATGprによる解析ステップ
- (2) 抽出された3つのORF上における、DNA配列のコーディング領域らしさの指標であるコーディングポテンシャル解析ステップ
- (3) アミノ酸配列データベースに対するホモロジー検索プログラムによる検索ステップ
- (4) 上記3つの解析結果をアミノ酸フレーム情報と同時に表示するステップ
- (5) 上記表示結果を参照しながら、フレームシフトエラーの可能性のある部分を編集するステップ
- (6) 上記解析結果と編集結果をハードディスクに保存するステップ

から構成される方法によって、信頼度の高いアミノ酸配列をcDNA配列から効率良く抽出するためのものであり、

本発明は、cDNA配列を入力する入力手段と、該入力されたcDNA配列を1文字ずつずらして翻訳して3つのアミノ酸フレームを得る翻訳手段と、前記入力されたcDNA配列とデータベース中のDNA配列又はアミノ酸配列とのアラインメントを生成し、そのアラインメントから類似性情報に基づいて前記入力されたcDNA配列から翻訳されるアミノ酸配列を決定するアラインメント手段と、前記3つのアミノ酸フレーム上にまたがって前記アラインメント手段によって決定されたアミノ酸配列の領域を線分で表示する表示手段とを備えるアミノ酸フレーム表示システムによって実現される。

【 0 0 0 8 】

また、本発明は、cDNA配列を入力する入力手段と、該入力されたcDNA配列を1文字ずつずらして翻訳して3つのアミノ酸フレームを得る翻訳手段と、該3つのアミノ酸フレームにおける各開始コドン及び終始コドンを予測するコドン予測手段と、前記3つのアミノ酸フレーム上に前記開始コドン及び終始コドンの位置を表示すると共に、開始コドン位置に開始コドンらしさを表現する量又は記号を表示する表示手段とを備えるアミノ酸フレーム表示システムによって実現される。

## 【 0 0 0 9 】

また、本発明は、cDNA配列を入力する入力手段と、該入力されたcDNA配列を1文字ずつずらして翻訳して3つのアミノ酸フレームを得る翻訳手段と、該3つのアミノ酸フレームにおける各開始コドン及び終始コドンを予測するコドン予測手段と、前記3つのアミノ酸フレームそれぞれにおけるコーディング領域らしさを表すコーディングポテンシャルを計算するコーディングポテンシャル計算手段と、前記3つのアミノ酸フレーム上に前記開始コドン及び終始コドンの位置を表示すると共に、各フレーム上又は別ウィンドーに、前記3つのアミノ酸フレームのコーディングポテンシャルを表示する表示手段とを備えるアミノ酸フレーム表示システムによって実現される。

## 【 0 0 1 0 】

また、本発明は、cDNA配列を入力する入力ステップと、該入力されたcDNA配列を1文字ずつずらして翻訳して3つのアミノ酸フレームを得る翻訳ステップと、前記入力されたcDNA配列とデータベース中のDNA配列又はアミノ酸配列とのアラインメントを生成し、そのアラインメントから類似性情報に基づいて前記入力されたcDNA配列から翻訳されるアミノ酸配列を決定するアラインメントステップと、前記3つのアミノ酸フレーム上にまたがって前記アラインメントステップによって決定されたアミノ酸配列の領域を線分で表示する表示ステップとを備えるアミノ酸フレーム表示方法によって実現される。

## 【 0 0 1 1 】

また、本発明は、cDNA配列を入力する入力ステップと、該入力されたcDNA配列を1文字ずつずらして翻訳して3つのアミノ酸フレームを得る翻訳ステップと、該3つのアミノ酸フレームにおける各開始コドン及び終始コドンを予測するコドン予測ステップと、前記3つのアミノ酸フレーム上に前記開始コドン及び終始コドンの位置を表示すると共に、開始コドン位置に開始コドンらしさを表現する量又は記号を表示する表示ステップとを備えるアミノ酸フレーム表示方法によって実現される。

## 【 0 0 1 2 】

また、本発明は、cDNA配列を入力する入力ステップと、該入力されたcD

N A 配列を 1 文字ずつずらして翻訳して 3 つのアミノ酸フレームを得る翻訳ステップと、該 3 つのアミノ酸フレームにおける各開始コドン及び終始コドンを予測するコドン予測ステップと、前記 3 つのアミノ酸フレームそれぞれにおけるコーディング領域らしさを表すコーディングポテンシャルを計算するコーディングポテンシャル計算ステップと、前記 3 つのアミノ酸フレーム上に前記開始コドン及び終始コドンの位置を表示すると共に、各フレーム上又は別ウィンドーに、前記 3 つのアミノ酸フレームのコーディングポテンシャルを表示する表示ステップとを備えることを特徴とするアミノ酸フレーム表示方法によって実現される。

## 【 0 0 1 3 】

また、本発明は、コンピュータを、c D N A 配列を入力する入力手段と、該入力された c D N A 配列を 1 文字ずつずらして翻訳して 3 つのアミノ酸フレームを得る翻訳手段と、前記入力された c D N A 配列とデータベース中の D N A 配列又はアミノ酸配列とのアラインメントを生成し、そのアラインメントによって決定される、前記入力された c D N A 配列から翻訳されたアミノ酸配列を得るアラインメント手段と、前記 3 つのアミノ酸フレーム上にまたがって前記アラインメント手段によって得られたアミノ酸配列の領域を線分で表示する表示手段とを備えるアミノ酸フレーム表示システムとして機能させるためのプログラムを記録したコンピュータ読み取り可能な記録媒体によって実現される。

## 【 0 0 1 4 】

また、本発明は、コンピュータを、c D N A 配列を入力する入力手段と、該入力された c D N A 配列を 1 文字ずつずらして翻訳して 3 つのアミノ酸フレームを得る翻訳手段と、該 3 つのアミノ酸フレームにおける各開始コドン及び終始コドンを予測するコドン予測手段と、前記 3 つのアミノ酸フレーム上に前記開始コドン及び終始コドンの位置を表示すると共に、開始コドン位置に開始コドンらしさを表現する量又は記号を表示する表示手段とを備えるアミノ酸フレーム表示システムとして機能させるためのプログラムを記録したことを特徴とするコンピュータ読み取り可能な記録媒体によって実現される。

## 【 0 0 1 5 】

また、本発明は、コンピュータを、c D N A 配列を入力する入力手段と、該入

力された cDNA 配列を 1 文字ずつずらして翻訳して 3 つのアミノ酸フレームを得る翻訳手段と、該 3 つのアミノ酸フレームにおける各開始コドン及び終始コドンを予測するコドン予測手段と、前記 3 つのアミノ酸フレームそれぞれにおけるコーディング領域らしさを表すコーディングポテンシャルを計算するコーディングポテンシャル計算手段と、前記 3 つのアミノ酸フレーム上に前記開始コドン及び終始コドンの位置を表示すると共に、各フレーム上又は別ウィンドーに、前記 3 つのアミノ酸フレームのコーディングポテンシャルを表示する表示手段とを備えるアミノ酸フレーム表示システムとして機能させるためのプログラムを記録したコンピュータ読み取り可能な記録媒体によって実現される。

## 【 0 0 1 6 】

## 【発明の実施の形態】

以下、添付図面を参照しながら本発明の好適な実施の形態について詳細に説明する。

図 1 は、本発明の一実施の形態によるアミノ酸フレーム表示システムの構成を示す図である。本実施の形態はディスプレイ 1、キーボード 2、中央演算装置 CPU 3、フロッピーディスク 5 が挿入されるフロッピーディスクドライブ 4、主メモリ 6、及び遺伝子配列データベース 7 から構成される。主メモリ 6 には、アミノ酸フレーム表示システムを実現するアミノ酸フレーム表示プログラムが格納され、そのアミノ酸フレーム表示プログラムは、入力手段 11、翻訳手段 12、アラインメント手段 13、表示手段 14、コドン予測手段 15、及び編集手段 16 の各手段に相当する機能を有する。このプログラムはディスプレイ 1、キーボード 2、フロッピーディスクドライブ 4、主メモリ 6、及び遺伝子配列データベース 7 などと共同して、CPU 3 で実行される。

## 【 0 0 1 7 】

次に、図 2 を用いてシステムの概要を説明する。ユーザー 101 によるシステムの立ち上げと共に、cDNA 配列及び解析パラメータの入力画面 102 が表示される。画面 102 中には、パラメータ表示及び配列表示の領域 103 にデフォルトのパラメータの値と cDNA 配列の入力ボックスが表示される。ユーザー 101 は、cDNA 配列の入力、及び解析パラメータの入力を行うことができ

る。画面 1 0 2 中には、解析処理を開始する解析ボタン 1 0 4 が表示されており、このボタンをユーザー 1 0 1 が押すことによって、cDNA 配列の解析及び表示処理が実行される。また、画面 1 0 2 中には、cDNA 配列解析結果を保存するハードディスク 1 1 8 中からの読み込み処理を開始する解析結果読み込みボタン 1 0 5 が表示されており、このボタンをユーザー 1 0 1 が押すことによって、cDNA 配列解析結果の表示処理が実行される。cDNA 配列の解析・表示処理 1 0 6、又は cDNA 配列解析結果読み込み表示処理 1 0 7 によって解析結果の表示画面 1 0 8（図 4 で詳述する）が表示される。表示画面 1 0 8 中には、解析結果表示及びパラメータ表示領域 1 0 9 に、解析結果及び解析パラメータ値が表示される。また表示画面 1 0 8 中には、パラメータ変更ボタン 1 1 0、編集ボタン 1 1 1、保存ボタン 1 1 2 が表示される。ユーザー 1 は、表示画面 1 0 8 中の解析結果 1 0 9 を閲覧しながら、解析パラメータを変更した上で解析再実行が可能である。パラメータ変更ボタン 1 1 0 を押すことによって、パラメータ変更処理 1 1 3 が開始される。パラメータ変更処理 1 1 3 後、cDNA 配列の解析・表示処理ステップ 1 0 6 が再度実行される。ユーザー 1 0 1 は、表示画面 1 0 8 中の解析結果 1 0 9 を閲覧しながら、cDNA 配列の編集が可能である。編集ボタン 1 1 1 を押すことによって、cDNA 配列の編集画面 1 1 4（図 8 で詳述する）が起動される。cDNA 配列の編集画面 1 1 4 では、cDNA 配列とアミノ酸配列とのアラインメント 1 1 5 が表示される。ユーザー 1 0 1 は、解析結果の表示画面 1 0 8 を参照しながら、アラインメント 1 1 5 の表示の中で、直接 cDNA 配列の編集を実行することが可能である。編集が終了したら、再解析ボタン 1 1 6 を押すことによって、cDNA 配列の解析・表示処理 1 0 6 を再度開始することができる。その結果は、再び解析結果の表示画面 1 0 8 上に表示され、編集の効果を確認することができる。ユーザー 1 0 1 は、cDNA 配列とその解析結果に名称を付けて電子ファイルとしてハードディスクに保存することができる。保存ボタン 1 1 2 を押すことによって、cDNA 配列とその解析結果の保存処理ステップ 1 1 7 が開始され、ハードディスク 1 1 8 のファイル中に保存される。

#### 【 0 0 1 8 】

次に、cDNA 配列の解析ステップについて図 3 を用いて詳細に説明する。ま



ず、入力された cDNA 配列から、ORF 抽出処理ステップ 201 によって、ORF 情報、すなわち、開始コドンと終始コドン、及びそれらのフレーム情報を抽出する。次に、cDNA 配列の類似性解析処理ステップ 202 を実行する。ステップ 202 では、アミノ酸配列データベース 209 をターゲットデータベースとして、BLASTX 検索処理ステップ 203 を実行する。BLASTX によって得られたヒットリストから、類似性尺度、例えば E-value の低い順に一定の数のデータベース エントリを抽出する。それらのアミノ酸配列と cDNA 配列との間で、次に、TRANSQ 解析処理ステップ 204 を実行する。TRANSQ は、cDNA 配列を翻訳しながら、cDNA 配列中のフレームシフトを考慮しつつ、アミノ酸配列との間でアラインメントを生成する。これによって、cDNA 配列中のフレームシフトを考慮した上での、cDNA 配列の翻訳アミノ酸配列が得られる。得られたアラインメントから、アラインメント情報抽出処理ステップ 205 によって、フレームシフト情報、及び cDNA 配列の翻訳アミノ酸配列情報を抽出する。ここで、cDNA 配列の翻訳アミノ酸配列情報は、BLASTX から得られるものを利用することも可能である。次に、cDNA 配列の統計解析処理ステップ 206 を実行する。まず、ATGpr 解析処理ステップ 207 によって、cDNA 配列に含まれる各開始コドン ATG に対して、開始コドンらしさのスコアを計算する。ATGpr はヘリックス研究所によって開発されたプログラム (Salamov, A.A., et al., Assessing Protein Coding Region Integrity in cDNA Sequencing Projects, Bioinformatics, 14, 384, 1998) で、cDNA 配列の持つ統計的性質を用いて開始コドンらしさのスコアを計算する。次に、コーディングポテンシャル解析処理ステップ 208 によって、コーディングポテンシャル解析処理を実行する。コーディングポテンシャル解析処理では、cDNA 配列内の一定長配列ウィンドーにおいて、コード領域らしさを各フレーム毎に計算し、ウィンドーをスライドさせながら、逐次コード領域らしさを計算していく。コード領域らしさの指標は、6 文字程度の塩基文字列の頻度統計解析によって得られる。

#### 【0019】

次に、解析結果の表示画面について説明する。図 4 に解析結果の表示画面の概要を示す。解析結果の表示画面 301 は、アミノ酸フレーム表示領域 302 (図

5で詳述する)、コーディングポテンシャル表示領域303(図6で詳述する)、及びアミノ酸アラインメント表示領域304(図7で詳述する)から構成される。

#### 【0020】

それぞれの領域について詳細に説明する。アミノ酸フレーム表示領域302には、アミノ酸フレームと共に類似性情報を表示する。その詳細を図5を用いて説明する。cDNA配列のスケール401を座標として、3つのアミノ酸配列フレームを表示する。すなわち、cDNA配列の5'端から1塩基目を出発点としてアミノ酸配列に翻訳していったフレーム1(402)、cDNA配列の5'端から2塩基目を出発点としてアミノ酸配列に翻訳していったフレーム2(403)、及びcDNA配列の5'端から3塩基目を出発点としてアミノ酸配列に翻訳していったフレーム3(404)を線分で表示する。各フレーム上には、開始コドン(ATG)の位置405、終始コドンの位置406を縦線で表示する。また、各フレーム上の開始コドンから終始コドンまでの線分(ORF)の中で、最長の線分(フレーム内最長ORF)を407で示す横線で表示する。以上は、通常行われる表示方法である。通常は、全フレームで最長のORFをもっともらしいORFとしてそのアミノ酸配列をその後の解析の対象にする。比較的長いORFが複数のフレームにまたがって存在する場合は、それらのORFの間の領域にフレームシフトが存在する可能性がある。この例では、フレーム1とフレーム2のフレーム内最長ORF間の領域に、フレームシフトが存在する可能性がある。しかし、この情報だけでは、どこにフレームシフトがあるかを特定することはできない。そこで、本発明では、既知アミノ酸配列との間の類似性情報とcDNA配列の持つ統計的な情報を利用する。類似性情報としては、cDNA配列とアミノ酸配列とのアラインメントから決定されたアミノ酸配列を、各フレームにまたがって線分408として表示する。線分408は、フレーム1とフレーム2にまたがって表示されており、フレーム間の遷移が生じている点でフレームシフトが生じていることがわかる。cDNA配列の持つ統計的な情報として、各開始コドンの近傍に、ATGprの出力409を表示する。これによって、各開始コドンから始まるORFのもっともらしさが、その長さだけではなく数値として表示される。

## 【 0 0 2 1 】

次に、コーディングポテンシャル表示領域 3 0 3 には、c D N A 配列に沿って、コーディングポテンシャル情報を表示する。その詳細を図 6 を用いて説明する。c D N A 配列のスケール 5 0 1 を横軸として、コーディングポテンシャルをその座標 5 0 2 上に表示する。コーディングポテンシャルとしては、フレーム 1 のコーディングポテンシャルの値 5 0 3、フレーム 2 のコーディングポテンシャルの値 5 0 4、及びフレーム 3 のコーディングポテンシャルの値 5 0 5 を表示する。チェックボックス 5 0 6、チェックボックス 5 0 7、チェックボックス 5 0 8 によって、それぞれフレーム 1、フレーム 2、フレーム 3 のコーディングポテンシャルの表示の有無を決定することができる。コーディングポテンシャルの計算においては、前述したように、c D N A 配列内の一定長配列ウィンドーにおいて、コード領域らしさを各フレーム毎に計算し、ウィンドーをスライドさせながら、逐次コード領域らしさを計算していく。コード領域らしさの指標は、6 文字程度の塩基文字列の頻度統計解析によって得られる。図 6 に示されるように、コーディングポテンシャルの値が高い領域がフレーム 1 からフレーム 2 に、1 3 0 塩基長付近で遷移している。これは、1 3 0 塩基長付近で、フレームシフトが存在することを示唆している。このように、コーディングポテンシャルの値のフレーム間の遷移をみることによって、フレームシフトの存在とその位置を推定することが可能である。コーディングポテンシャルの計算時のウィンドーサイズとシフト値はボックス 5 0 9 及びボックス 5 1 0 内に表示される。これらのボックス内の値は、変更してコーディングポテンシャルを再計算して表示することが可能である。これは、ボタン 5 1 1 を押すことによって可能である。

## 【 0 0 2 2 】

次に、アミノ酸アラインメント表示領域 3 0 4 には、アミノ酸配列データベース中のアミノ酸配列との間のアラインメントを線分によって表示する。その詳細を図 7 を用いて説明する。c D N A 配列のスケール 6 0 1 を座標として、アミノ酸配列データベース中のアミノ酸配列との間のアラインメントを線分で表示する。アミノ酸データベースとしては、SWISS-PROT や OWL 等が用いられる。アラインメントは、図 3 の説明で記述したように、TRANSQ 又は BLASTX によって得られるア

ラインメントを用いる。ここでは、TRANSQを用いた場合の説明を行う。図3で説明したように、TRANSQの比較を行う前のBLASTXの検索によって得られたE-valueの値でソートしたアラインメントを、線分で表示する。アラインメントは、E-valueの値が小さい順に、上から下に向かって並べられる。この例として、cDNA配列とアミノ酸配列との第一のアラインメント602、cDNA配列とアミノ酸配列との第二のアラインメント603、及びcDNA配列とアミノ酸配列との第三のアラインメント604を示す。アラインメント602、603、及び604の左側には、各アラインメントを特徴付ける数値情報614 (Identity、blastx検索のE-value、アラインメント長 (Al)、5' 端のアラインメントされていないDNA側の長さ (Nab)、5' 端のアラインメントされていないアミノ酸側の長さ (Naa)) が記述される。アラインメント602、603、及び604の右側には、アミノ酸配列に関する情報615 (ID、Definition等) が記述される。各線分において、アラインメントされていない領域を線分608で表示する。アラインメントされた領域は、アラインメントのIdentityの値に応じて識別可能なパターンで表示される。この一致度は色によって表示しても良い。線分605は、Identity $\geq$ 90%の領域を、線分606は、90% $>$ Identity $\geq$ 40%の領域を、線分607は、40% $>$ Identityの領域をそれぞれ表す。Identityの値は一定サイズのウィンドー内で計算し、配列に沿ってスライドさせることによって配列の各領域における値を計算する。cDNA配列とアミノ酸配列とのアラインメント中において、DNA側の挿入 (挿入数が3の倍数の場合) 領域を、線分609で表し、DNA側の欠失 (欠失数が3の倍数の場合) 領域を線分610で表す。これによって、アラインメントから得られるアミノ酸配列中のフレームシフトの情報が、複数のアラインメントについて同時に確認可能になる。また、アラインメントの挿入欠失が生じている部位がどのIdentity領域であるかによって、それらの挿入欠失の有意性を判断することが可能である。すなわち、Identityが高い領域で挿入欠失が生じていれば、その有意性は高く、逆にIdentityが低い領域で挿入欠失が生じていれば、その有意性は低い。例えば、アラインメント602上と、アラインメント603上の同じ位置に存在するDNA側の挿入609は、その位置におけるIdentityが90%以上あることから、有意性が高いと判断できる。一方、ア

ラインメント603上に存在するDNA側の欠失610は、その位置におけるIdentityが40%以下であることから、有意性が低いと判断できる。また、アラインメント604をみると、このcDNAはリボソーム蛋白質とIdentity100%で相同性を示していることから、リボソーム遺伝子とキメラ遺伝子を構成しており、その接続部位は300塩基付近であることが推定される。ユーザーは複数のアラインメント上の挿入欠失部位とその部位のIdentityを総合的に観察することによって、cDNA配列の編集をすべき部位とそのためにどのアラインメントを用いるべきかを判断することが可能である。各アラインメントの詳細情報へのリンクは、アラインメント線分の右横のチェックボックス611、612、613によって行われる。例えば、チェックボックス611を選択することによって、図5で説明したアミノ酸フレーム表示上に、選択した一番目のアラインメントから得られたアミノ酸配列を線分として表示することが可能である。アミノ酸フレーム表示の項で説明したように、アラインメントから得られたアミノ酸配列の線分表示と、cDNA配列自体から得られるORFの線分表示とを、同時に3つのアミノ酸フレーム上で比較することによって、フレームシフトがどこで生じどのフレームからどのフレームへ遷移しているかが明確になる。こうやって、フレームシフトの存在と確からしさを図5と図7とで確認しながら、アラインメント中のcDNA配列のフレームシフト部位を編集することが可能である。cDNA配列の編集画面へのリンクは、編集したいアラインメントを、チェックボックス611、612、又は613によって選択する。そして、図2で示した編集ボタン111を押すことによって、編集画面114が生成される。

#### 【0023】

図8に編集画面の詳細を示す。編集画面には、cDNA配列とアミノ酸配列とのアラインメント701と編集のための各種ボタン703、704、705が表示される。cDNA配列とアミノ酸配列とのアラインメント701では、cDNA配列とその翻訳アミノ酸配列がターゲットの既知アミノ酸配列と並置されてテキスト表示される。翻訳アミノ酸配列とターゲットのアミノ酸配列との間の実線は、アミノ酸が一致していることを表し、二つの点と一つの点は点の数に応じたアミノ酸間の類似度の大きさを表している。このアラインメントでは、702に

示す位置に a 塩基の挿入が生じていることがわかる。すなわち、a 塩基を挿入塩基とみなすことによって、a 塩基の前後のアミノ酸配列が良く一致することがわかる。編集は、この a 塩基を直接ユーザーが消去することによって行われる。編集結果確定と再解析の実行は、cDNA 配列の編集確定再解析ボタン 703 (Submit) を押すことによって可能である。これによって、cDNA 配列の解析・表示処理 106 が再度実行される。編集画面の終了は、cDNA 配列とアミノ酸配列とのアラインメント、及び編集画面の終了ボタン 704 (Close) を押すことによって可能である。編集確定再解析の前に編集した結果のリセットは、cDNA 配列編集のリセットボタン 705 (Refresh) を押すことによって可能である。

## 【 0 0 2 4 】

こうやって cDNA 配列の編集後再解析を行った結果は、解析結果の表示画面 108 に直ちに反映される。図 8 で挿入塩基とみなされた a 塩基を消去する編集を行った結果を、アミノ酸フレーム表示について図 9 に示す。図 5 と比較することによって、130 塩基付近以上の領域で、各フレーム上の情報が入れ替わっていることがわかる。アラインメントによって決定されたアミノ酸配列のフレーム上での線分が、図 5 ではフレーム 1 とフレーム 2 の間でまたがっていたが、図 9 では一つの線分となって、フレーム 1 上に表示されていることがわかる。これによって、図 8 で行った cDNA 配列の編集の妥当性を確認することができる。また、ATGpr の値が、編集によって更新されていることがわかる。特にフレーム 1 上の左端の ATG についてのスコア値が、0.45 から 0.80 に増加しており、これは編集によって左端の ATG から始まる ORF が長くなったことに起因していると考えられる。このように ATGpr スコア値の増加によって、編集の妥当性をさらに確認することができる。

## 【 0 0 2 5 】

なお、本発明は上記実施の形態に限定されるものではない。

本発明は、コンピュータを上記アミノ酸フレーム表示システムとして機能させるためのプログラムを記録したコンピュータ読み取り可能な記録媒体であってもよく、例えば、磁気テープ、CD-ROM、IC カード、RAM カード等のいか

なるタイプの記録媒体であってもよい。

【 0 0 2 6 】

【発明の効果】

本発明によれば、未知 cDNA 配列の各アミノ酸フレーム上の ORF 表示に加えて、既知アミノ酸配列との類似性比較から得られた cDNA 配列のアミノ酸情報を該フレーム上に表現し、同時に統計的に得られた ORF に関する情報（開始コドンらしさ、及びコーディングポテンシャルグラフ）を表示することによって、効率的にフレームシフトを検出し、その結果を編集することによって、高精度なアミノ酸配列を得ることが可能となる。

【図面の簡単な説明】

【図 1】

本発明の一実施の形態によるアミノ酸フレーム表示システムの構成を示す図。

【図 2】

本発明の一実施の形態における、画面遷移と解析処理の流れを示す図。

【図 3】

配列解析のフローチャートを表した図。

【図 4】

解析結果の表示画面の概要を表した図。

【図 5】

アミノ酸フレーム上への類似性情報の表示方法（編集前の画面例）を表した図

。

【図 6】

cDNA 配列に沿ったコーディングポテンシャルの表示方法を表した図。

【図 7】

アミノ酸との類似性情報の可視化表示方法を表した図。

【図 8】

アラインメントのテキスト表示と編集画面を表した図。

【図 9】

アミノ酸フレーム上への類似性情報の表示方法（編集後の画面例）を表した図

【符号の説明】

- 1 0 1 本システムを操作するユーザ
- 1 0 2 cDNA配列、及び解析処理パラメータの入力画面
- 1 0 3 パラメータ表示及び配列表示の領域
- 1 0 4 解析処理を開始する解析ボタン
- 1 0 5 解析結果の読み込みと再解析処理を開始するボタン
- 1 0 6 cDNA配列の解析処理及び表示処理ステップ
- 1 0 7 cDNA配列解析結果の読み込み処理及び表示処理ステップ
- 1 0 8 解析処理結果の表示画面
- 1 0 9 解析処理結果表示及びパラメータ表示の領域
- 1 1 0 パラメータ変更処理を開始するボタン
- 1 1 1 編集画面を起動するボタン
- 1 1 2 解析処理結果の保存処理を開始するボタン
- 1 1 3 パラメータ変更処理ステップ
- 1 1 4 cDNA配列の編集画面
- 1 1 5 アラインメント表示の領域
- 1 1 6 再解析処理を開始するボタン
- 1 1 7 cDNA配列と解析結果の保存処理ステップ
- 1 1 8 cDNA配列及び解析結果が保存されたハードディスク
- 2 0 1 cDNA配列のORF抽出処理ステップ
- 2 0 2 cDNA配列の類似性解析処理ステップ
- 2 0 3 BLASTX解析処理ステップ
- 2 0 4 TRANSQ解析処理ステップ
- 2 0 5 アラインメント情報抽出処理ステップ
- 2 0 6 cDNA配列の統計解析処理ステップ
- 2 0 7 ATGpr解析処理ステップ
- 2 0 8 コーディングポテンシャル解析処理ステップ
- 2 0 9 アミノ酸配列データベース



- 3 0 1 解析結果の表示画面
- 3 0 2 アミノ酸フレーム表示領域
- 3 0 3 コーディングポテンシャル表示領域
- 3 0 4 アミノ酸アラインメント表示領域
- 4 0 1 cDNA配列のスケール
- 4 0 2 cDNA配列の5' 端から1塩基目を出発点としてアミノ酸配列に翻訳していったフレーム1
- 4 0 3 cDNA配列の5' 端から2塩基目を出発点としてアミノ酸配列に翻訳していったフレーム2
- 4 0 4 cDNA配列の5' 端から3塩基目を出発点としてアミノ酸配列に翻訳していったフレーム3
- 4 0 5 各フレーム上の開始コドン (A T G) の位置
- 4 0 6 各フレーム上の終始コドンの位置
- 4 0 7 各フレーム上の開始コドンから終始コドンまでの線分 (O R F) の内で最長の線分 (フレーム内最長O R F)
- 4 0 8 cDNA配列とアミノ酸配列とのアラインメントから決定されたアミノ酸配列を各フレームにまたがって表示した線分
- 4 0 9 各開始コドンについて、開始コドンから始まるO R Fのもっともらしさを表示する量 (ATGprの出力)
- 5 0 1 cDNA配列のスケール
- 5 0 2 cDNA配列に沿った各領域におけるコーディングポテンシャルの値を示す座標
- 5 0 3 フレーム1のコーディングポテンシャルの値
- 5 0 4 フレーム2のコーディングポテンシャルの値
- 5 0 5 フレーム3のコーディングポテンシャルの値
- 5 0 6 フレーム1のコーディングポテンシャルについて表示の有無を決定するチェックボックス
- 5 0 7 フレーム2のコーディングポテンシャルについて表示の有無を決定するチェックボックス

508 フレーム3のコーディングポテンシャルについて表示の有無を決定するチェックボックス

509 コーディングポテンシャルの値を計算するウィンドーサイズの表示、及び変更値を入力するボックス

510 コーディングポテンシャルの値を計算するウィンドーサイズのシフト値の表示、及び変更値を入力するボックス

511 コーディングポテンシャルの値を計算するウィンドーサイズ、及びシフト値を変更して再計算するためのボタン

601 cDNA配列のスケール

602 cDNA配列とアミノ酸配列との第一のアラインメント

603 cDNA配列とアミノ酸配列との第二のアラインメント

604 cDNA配列とアミノ酸配列との第三のアラインメント

605 cDNA配列とアミノ酸配列とのアラインメント中において、Identity $\geq$ 90%の領域

606 cDNA配列とアミノ酸配列とのアラインメント中において、90%>Identity $\geq$ 40%の領域

607 cDNA配列とアミノ酸配列とのアラインメント中において、40%>Identityの領域

608 cDNA配列とアミノ酸配列とのアラインメント中において、アラインメントされていない領域

609 cDNA配列とアミノ酸配列とのアラインメント中において、DNA側の挿入（挿入数が3の倍数の場合）領域

610 cDNA配列とアミノ酸配列とのアラインメント中において、DNA側の欠失（欠失数が3の倍数の場合）領域

611 cDNA配列とアミノ酸配列との第一のアラインメントを選択するチェックボックス

612 cDNA配列とアミノ酸配列との第二のアラインメントを選択するチェックボックス

613 cDNA配列とアミノ酸配列との第三のアラインメントを選択するチ

エックボックス

6 1 4 cDNA配列とアミノ酸配列とのアラインメントの特徴を表す数値 (Identity、blastx検索のE-value、アラインメント長、5' 端のアラインメントされていないDNA側の長さ、5' 端のアラインメントされていないアミノ酸側の長さ) の表示領域

6 1 5 cDNA配列とアミノ酸配列とのアラインメントについて、アミノ酸に関する情報 (ID、Definition等)

7 0 1 cDNA配列とアミノ酸配列とのアラインメント表示

7 0 2 cDNA配列中のa塩基の挿入の例

7 0 3 cDNA配列の編集確定ボタン

7 0 4 cDNA配列とアミノ酸配列とのアラインメント、及び編集画面の終了ボタン

7 0 5 cDNA配列の編集のリセットボタン

8 0 1 cDNA配列のスケール

8 0 2 cDNA配列の5' 端から1塩基目を出発点としてアミノ酸配列に翻訳していったフレーム1

8 0 3 cDNA配列の5' 端から2塩基目を出発点としてアミノ酸配列に翻訳していったフレーム2

8 0 4 cDNA配列の5' 端から3塩基目を出発点としてアミノ酸配列に翻訳していったフレーム3

8 0 5 各フレーム上の開始コドン (ATG) の位置

8 0 6 各フレーム上の終始コドンの位置

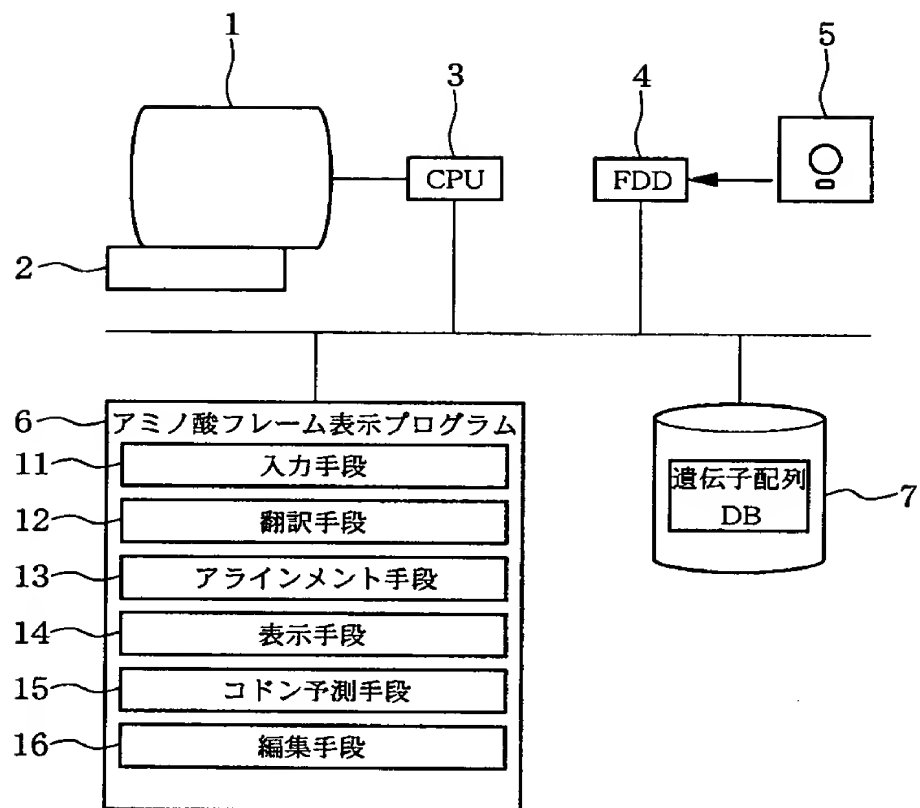
8 0 7 各フレーム上の開始コドンから終始コドンまでの線分 (ORF) の内で最長の線分 (フレーム内最長ORF)

8 0 8 cDNA配列とアミノ酸配列とのアラインメントから決定されたアミノ酸配列を各フレームにまたがって表示した線分

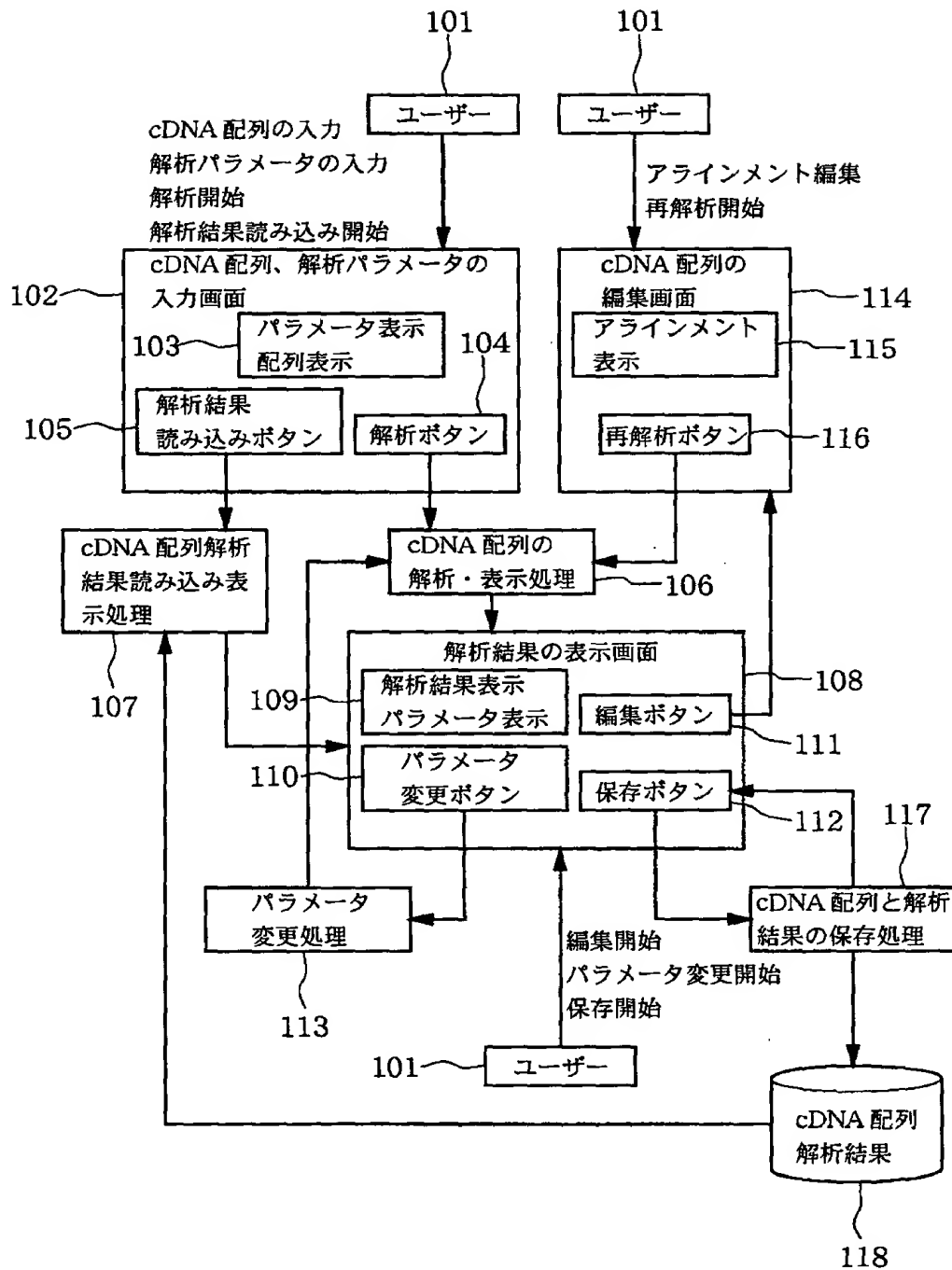
8 0 9 各開始コドンについて、開始コドンから始まるORFのもっともらしさを表示する量 (ATGprの出力)

【書類名】 図面

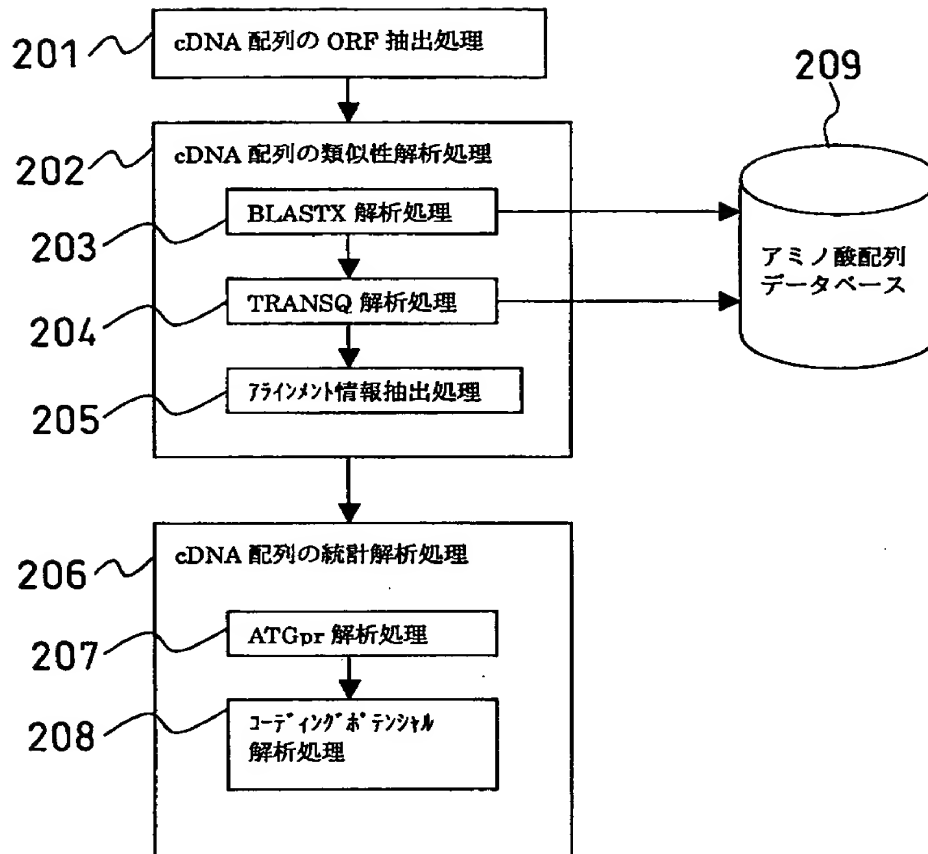
【図 1】



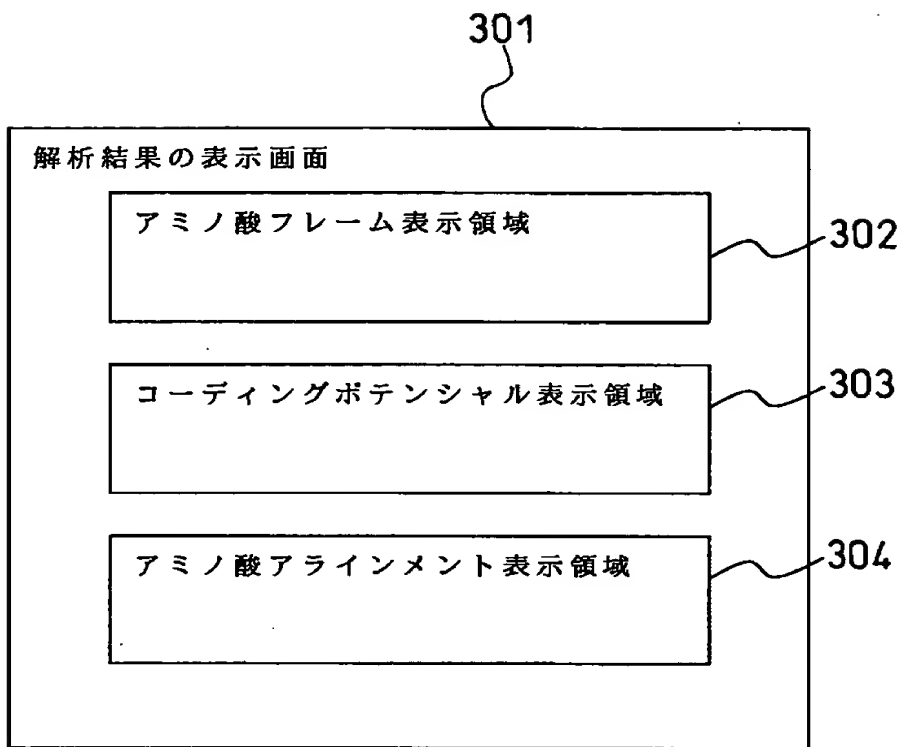
【図 2】



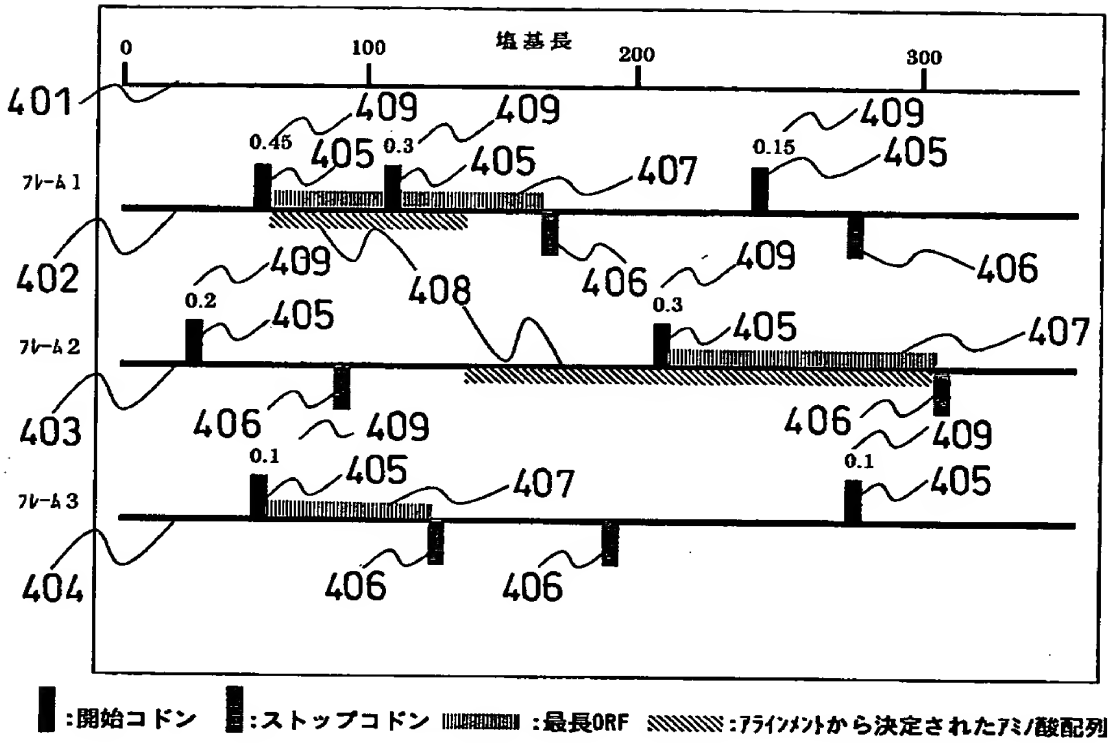
【図 3】



【図 4】

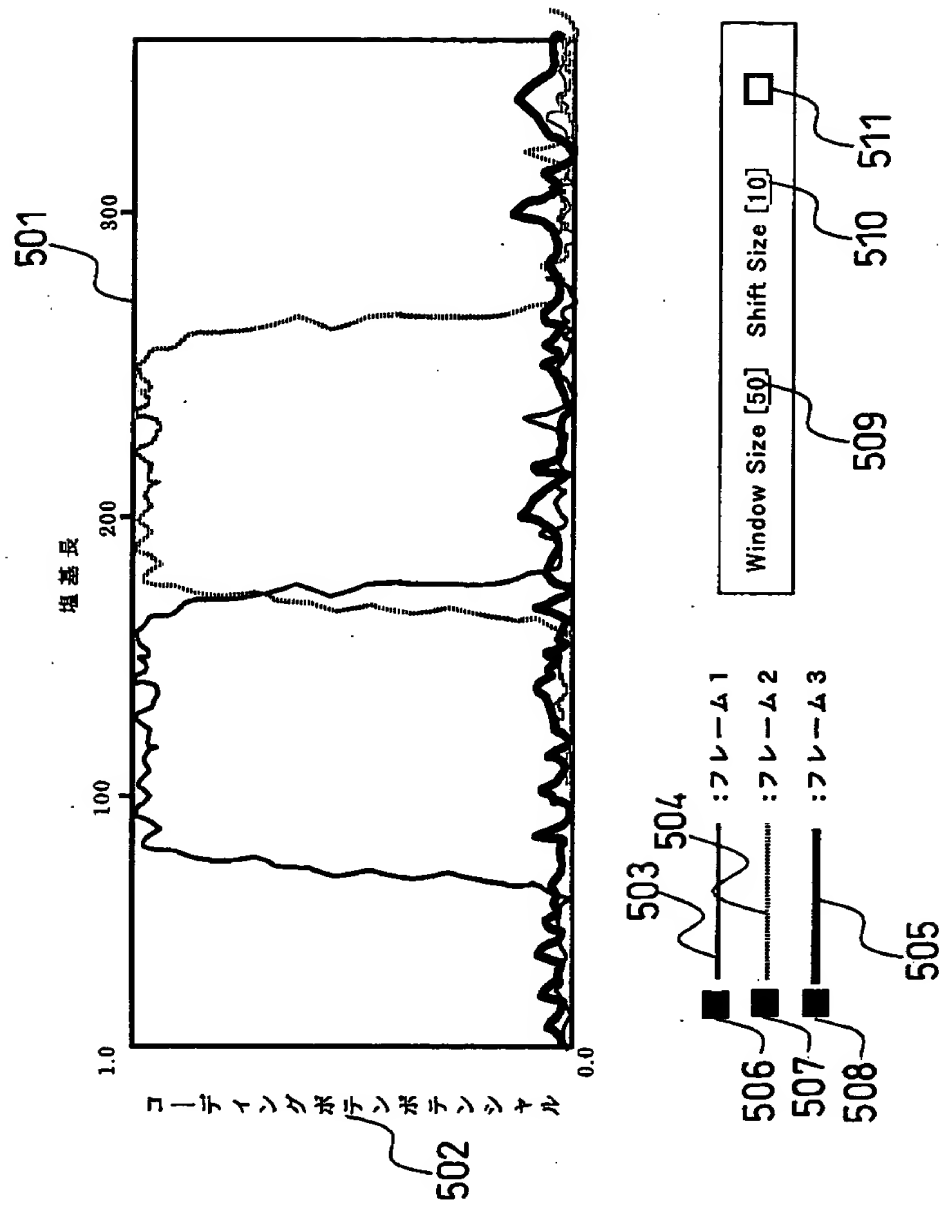


【図 5】

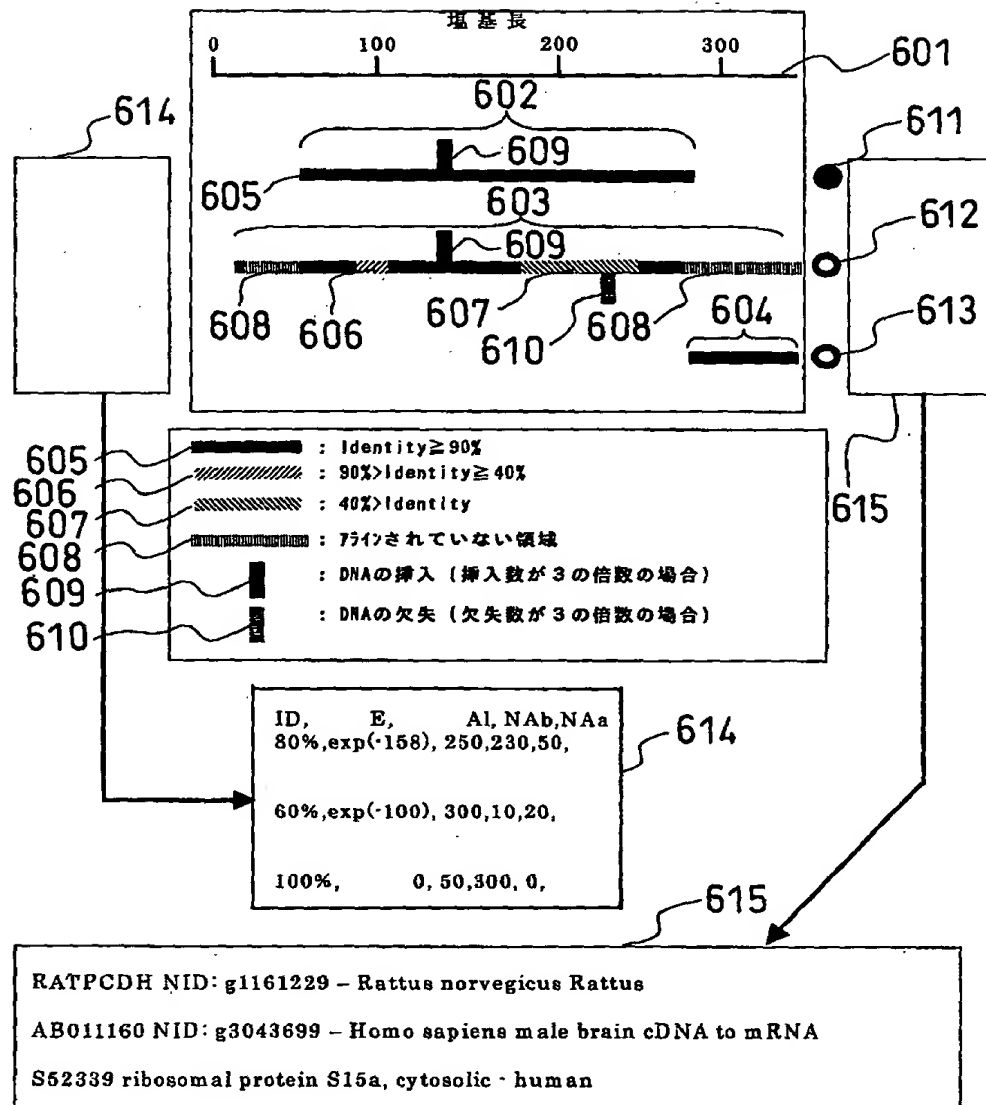




【図 6】



【図7】



【図 8】

701

Query sequence:>C-HEMBA1001174

Target sequence:>|c|ARL5\_RAT ADP-RIBOSYLATION FACTOR-LIKE PROTEIN

5. - RATTUS NORVEGICUS (RAT)

Score: 794 + strand

Alignment region Query : 208 .. 744

Target: 1 .. 179

Identity: 144/179(80%) Positive: 165/179(92%)

60 :	70 :	80 :	90 :	100 :	110 :
Query: 208 ATGgggctgatcttcgccaaactgtggagcctcttctgtaaccaagaacacaaagtaatt					
MetGlyLeuIlePheAlaLysLeuTrpSerLeuPheCysAsnGlnGluHisLysValIle					
: :   . : :       :					
Target: 1 MetGlyIleLeuPheThrArgIleTrpArgLeuPheAsnHisGlnGluHisLysValIle					

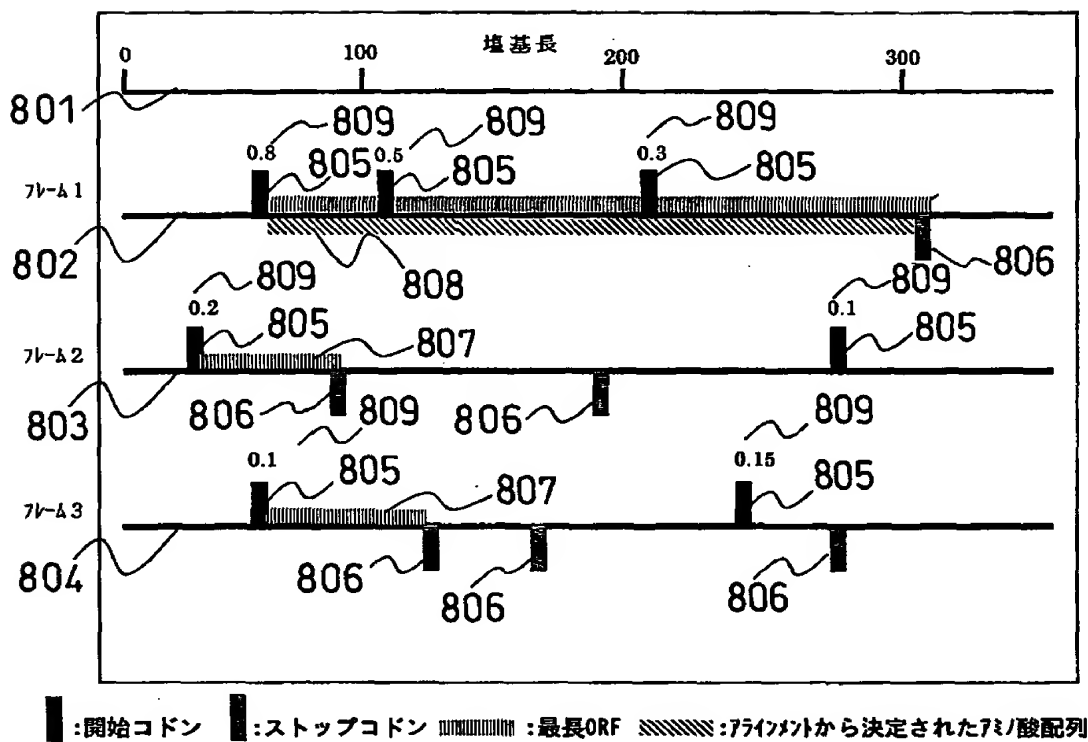
  

120 :	130 :	140 :	150 :	160 :	170 :
Query: 267 atagtgggactgagataatgcagggaaaaccaccattctttaccaattcttaATGaatgaa					
IleValGlyLeu AspAsnAlaGlyLysThrThrIleLeuTyrGlnPheLeuMetAsnGlu					
Target: 20 IleValGlyLeu AspAsnAlaGlyLysThrThrIleLeuTyrGlnPheSerMetAsnGlu					

702

Submit	Close	Refresh
703	704	705

【図9】



【書類名】 要約書

【要約】

【課題】 cDNA配列にフレームシフトエラーが存在する場合にも、信頼度の高いアミノ酸配列をcDNA配列から効率良く抽出することができるアミノ酸フレーム表示システム、アミノ酸フレーム表示方法及び記録媒体を提供すること。

【解決手段】 cDNA配列の各アミノ酸フレーム上のORF表示に加えて、既知アミノ酸配列との類似性比較から得られたcDNA配列のアミノ酸情報を該フレーム上に表現し、同時に統計的に得られたORFに関する情報として開始コドンらしさ、及びコーディングポテンシャルグラフを表示することによって、効率的にフレームシフトを検出し、その結果を編集することによって、高精度なアミノ酸配列を得ることが可能となる。

【選択図】 図2

【書類名】 手続補正書  
【整理番号】 H001696  
【提出日】 平成12年10月31日  
【あて先】 特許庁長官 殿  
【事件の表示】  
【出願番号】 特願2000-325403  
【補正をする者】  
【識別番号】 597059742  
【氏名又は名称】 株式会社ヘリックス研究所  
【代理人】  
【識別番号】 100091096  
【弁理士】  
【氏名又は名称】 平木 祐輔  
【手続補正 1】  
【補正対象書類名】 特許願  
【補正対象項目名】 提出物件の目録  
【補正方法】 追加  
【補正の内容】  
【提出物件の目録】  
【物件名】 委任状 1

(A)10002070316  
10002070316

## 委任状

平成12年10月24日

千葉県木更津市矢那1532番地3

株式会社 ヘリックス研究所

取締役研究所長 増保 安彦



私は、識別番号100091096（弁理士）平木 祐輔氏を以って代理人として下記事項を委任します。

記

特願2000-325403

1. 特許出願（「遺伝子配列情報の表示方法及び記録媒体」）に関する一切の件
2. 上記出願に関する放棄又は取下げ
3. 上記出願に関する出願審査等の請求又は申立ての取下げ
4. 上記出願に関する優先審査等の申請又は申立ての取下げ
5. 上記出願に関する出願人名義変更
6. 上記出願に基づく特許法第41条第1項の優先権主張並びにその取下げ
7. 上記出願に基づく実用新案法第10条第1項の出願変更
8. 上記出願に基づく特許法第121条第1項の審判の請求並びにその取下げ
9. 上記出願に基づくすべての特許権の存続期間の延長登録の出願に関する一切の件及びその出願取下げ
10. 上記1項を処理するための復代理人の選任及び解任

以上

特 2000-325403

**認定・付加情報**

特許出願の番号	特願 2000-325403
受付番号	10002070316
書類名	手続補正書
担当官	仲村 百合子 1730
作成日	平成13年 2月19日

**<認定情報・付加情報>**

**【提出された物件の記事】**

【提出物件名】	委任状（代理権を証明する書面）	1
---------	-----------------	---



【書類名】 手続補正書

【整理番号】 H001696

【提出日】 平成13年 1月11日

【あて先】 特許庁長官 殿

【事件の表示】

【出願番号】 特願2000-325403

【補正をする者】

【識別番号】 597059742

【氏名又は名称】 株式会社ヘリックス研究所

【代理人】

【識別番号】 100091096

【弁理士】

【氏名又は名称】 平木 祐輔

【発送番号】 086530

【手続補正 1】

【補正対象書類名】 特許願

【補正対象項目名】 提出物件の目録

【補正方法】 追加

【補正の内容】

【提出物件の目録】

【物件名】 委任状 1

【その他】 平成13年1月11日付にて印鑑変更届を提出。

(A)10100060049  
7

## 委任状

平成13年1月4日

千葉県木更津市矢那1532番地3

株式会社 ヘリックス研究所

代表取締役社長 永山 海



私は、識別番号100091096（弁理士）平木 祐輔 氏を以って代理人として下記事項を委任します。

### 記

1. 特願2000-325403に関する一切の件
2. 上記出願に関する放棄又は取下げ
3. 上記出願に関する出願審査等の請求又は申立ての取下げ
4. 上記出願に関する優先審査等の申請又は申立ての取下げ
5. 上記出願に関する出願人名義変更
6. 上記出願に基づく特許法第41条第1項の優先権主張並びにその取下げ
7. 上記出願に基づく実用新案法第10条第1項の出願変更
8. 上記出願に基づく特許法第121条第1項の審判の請求並びにその取下げ
9. 上記出願に基づくすべての特許権の存続期間の延長登録の出願に関する一切の件及びその出願取下げ
10. 上記1項を処理するための復代理人の選任及び解任

以上

特 2000-325403

## 認定・付加情報

特許出願の番号	特願2000-325403
受付番号	10100060049
書類名	手続補正書
担当官	仲村 百合子 1730
作成日	平成13年 2月19日

### <認定情報・付加情報>

#### 【提出された物件の記事】

【提出物件名】	委任状（代理権を証明する書面）	1
---------	-----------------	---

次頁無

特2000-325403

出願人履歴情報

識別番号 [000005108]

1. 変更年月日 1990年 8月31日

[変更理由] 新規登録

住 所 東京都千代田区神田駿河台4丁目6番地

氏 名 株式会社日立製作所

特2000-325403

出 願 人 履 歴 情 報

識別番号

[597059742]

1. 変更年月日

1997年 4月28日

[変更理由]

新規登録

住 所

千葉県木更津市矢那1532番地3

氏 名

株式会社ヘリックス研究所